International Conference on Control, Automation and Systems 2008
Oct. 14-17, 2008 in COEX, Seoul, Korea

Stable Monocular SLAM with Indistinguishable Features on Estimated Ceiling Plane Using Upward Camera

Seo-Yeon Hwang and Jae-Bok Song

Department of Mechanical Engineering, Korea University, Seoul, Korea
(Tel : +82-2-3290-3363; E-mail: etoile02, jbsong@korea.ac.kr)

Abstract: This paper deals with monocular SLAM of a mobile robot in an indoor environment where visually similar features (so-called indistinguishable features) exist on the ceiling. When these indistinguishable features exist near a feature, data association suffers from false matches which lead to localization failure. To reliably estimate the robot pose in various environments, the proposed scheme uses the additional information on whether the features are easily distinguishable (unique) or not. Then, a ceiling plane is estimated by the heights of the unique features to find the positions of the indistinguishable features approximately. Reliable data association is done by registering the indistinguishable features on the ceiling plane with small uncertainty regions which prevent the false matches. After this preprocess step, the data association results are used within the extended Kalman filter to estimate the robot pose. Corners and lamps are used as features in the experiments, and the results show that the proposed method successfully works in various indoor environments including indistinguishable features.

Keywords: Mobile robot, monocular SLAM, upward camera, ceiling.

1. INTRODUCTION

Simultaneous localization and mapping (SLAM) is one of the most important issues in mobile robotics. Since a mobile robot in the unknown environment has no initial information, localization and mapping should be conducted simultaneously to accomplish the given navigation task. In general, a mobile robot uses several sensors to extract features for SLAM, and the feature extraction methodologies are different for each sensor. The sensors used in SLAM can be divided into range sensors and vision sensors. Range sensors provide the range information directly, so they do not require a complicated process to extract features. However, range sensors such as laser scanners, sonar sensors, and IR scanners are expensive, and are usually used for expensive robots. Vision sensors require complex processes to extract features, but they have drawn much attention in recent years because of their low cost.

In recent years, several SLAM approaches have adopted a monocular camera and successful performance was demonstrated in [1, 2] with corner and line features. Consistent tracking of the feature is an important factor for the accuracy of localization in the monocular SLAM. If the number of extracted features is not enough and this situation lasts for a long time, it increases the pose uncertainty of a camera and finally leads to localization failure. Since the environment does not always have one kind of feature, the use of various types of features (e.g., corner, line, lamp, etc.) provides better tracking performance than that of a single feature type, and makes it possible that the SLAM algorithm can be applied to various environments and situations.

In mobile robotics, several approaches to combining odometric and visual information have been proposed. For example, an upward camera based SLAM using corner and line features demonstrated good performance [3, 4], and was successfully applied to the robot vacuum cleaners. This scheme has several advantages when the robot uses the features obtained from the ceiling image for localization. First, it is not affected by dynamic obstacles such as moving people, which results in consistent tracking, while a forward-looking camera suffers from dynamic obstacles substantially. Second, there is only a small variation in scale of features between successive ceiling images. Therefore, the whole scheme of the SLAM algorithm takes less time since an extra algorithm which considers the scale variation of features is not required.

However, despite the advantages of using the ceiling images mentioned above, data association in the feature tracking process can fail due to indistinguishable features. These types of features can be easily found in the cross-striped ceiling at the office, coffered ceiling at home, and even in the reflectors, as shown in Fig 1. If these features are densely distributed in the ceiling image, false matches occur during the data association process. Then, the uncertainty regions of these features in the feature map are likely to converge to a wrong position, which causes the localization failure since the robot pose is related to the feature positions. This is one of the critical problems that should be solved to accomplish stable navigation in various indoor environments.

Fig. 1 Environments where indistinguishable features can be easily extracted; (a) cross-striped ceiling, (b) coffered ceiling, and (c) reflector.

Usually, the outliers such as false matched features are successfully removed by the RANSAC (Random Sample Consensus) algorithm [5]. RANSAC provides a good result when the number of inliers is large enough compared to that of outliers. However, in the case of the monocular SLAM using an upward camera, the algorithm uses a minimized number of features to maintain the real-time performance in a large environment. Thus, in this case, RANSAC is not an appropriate method to deal with the indistinguishable features. To cope with this problem, a specific method which is optimized to this SLAM scheme is required.

This paper proposes a novel method for data association for the monocular SLAM using an upward camera. The corner and lamp features are combined to estimate the robot pose. The proposed algorithm determines whether the newly extracted features are unique (easily distinguishable from nearby features) or not. A plane equation of the ceiling is estimated according to the heights of the unique features. The indistinguishable features which lie on this plane are then registered to the database with small uncertainty regions. Data associations of the indistinguishable features from two sequential images are conducted reliably by the small matching boundaries acquired from the uncertainty regions.

The rest of this paper is organized as follows. In section 2, the procedures of the EKF (Extended Kalman Filter) based the monocular SLAM using corner and lamp features are introduced. Section 3 presents the details of the feature extraction methods. Section 4 describes the reliable data association with indistinguishable corner features. Results from the real environment are presented in section 5 and conclusions are drawn in section 6.

2. EKF-BASED SLAM

The basic concept of the monocular SLAM is introduced in [6]. Since a monocular camera cannot provide the distance to the features directly, the robot should observe the features at different positions to reduce the position uncertainty. The most efficient way to reduce the feature uncertainty in a short time is to make observations at substantially different angles. Thus, the robot equipped with the camera looking on the left or right side [7] or the upper direction can give the best results.

The EKF (Extended Kalman Filter) algorithm [8] is adopted to handle nonlinearities involved in the robot motion. Since the corner and lamp features are combined to estimate the robot pose, the following state vector and covariance matrix are defined.

$$X = [X_R^T, X_{C_1}^T, \cdots, X_{C_n}^T, X_{L_1}^T, \cdots, X_{L_m}^T]^T \tag{1}$$

$$P = \begin{bmatrix} P_R & P_{RC} & P_{RL} \\ P_{CR} & P_C & P_{CL} \\ P_{LR} & P_{LC} & P_L \end{bmatrix} \tag{2}$$

where $X_R = [x_R, y_R, \theta_R]^T$ is the position (x_R and y_R) and the orientation (θ_R) of the robot, $X_{C_i} = [x_{C_i}, y_{C_i}, z_{C_i}]^T$ and $X_{L_j} = [x_{L_j}, y_{L_j}, z_{L_j}]^T$ denote the positions of the i-th corner and j-th lamp, as shown in Fig. 2, where n and m are the total number of corners and lamps, respectively. P_R, P_C, and P_L represent the covariance matrices of the robot, all corners and lamps, respectively, and P_{RC} (and P_{CR}), P_{CL} (and P_{LC}), and P_{RL} (and P_{LR}) are the cross-covariance matrices related to P_R and P_C, P_C and P_L, and P_R and P_L, respectively. The EKF algorithm consists of the prediction and update stages, and the details are represented in the following subsections.

2.1 Prediction

At the prediction stage, the state vector $\hat{X}_k$ and its covariance matrix P_k at time k are calculated from $\hat{X}_{k-1}$ and P_{k-1} at time k-1 and the encoder reading u_k as follows:

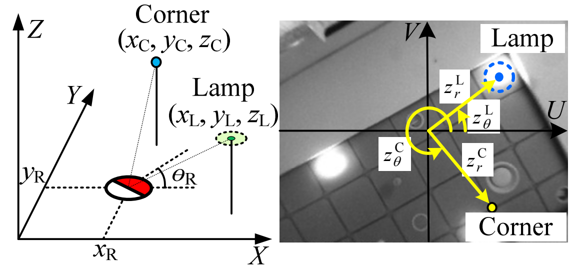


Fig. 2 Measurements in global and image coordinates.

$$\hat{X}_k^- = f(\hat{X}_{k-1}, u_k) \tag{3}$$

$$P_k^- = \nabla F_x P_{k-1} \nabla F_x^T + \nabla F_u Q \nabla F_u^T \tag{4}$$

where Q represents the covariance matrix of the process noise, f is a function of the system dynamics, and $\nabla F_x = \partial f / \partial X$ and $\nabla F_u = \partial f / \partial u$ are the Jacobian matrices of the nonlinear function f with respect to the state and input, respectively. Note that the superscript "-" indicates the state before the measurement at time k is taken.

If the robot sees the previously observed feature, the EKF performs the update stage. The prediction of a feature on the sensor frame, or the image coordinate, can be obtained by the following observation model based on the predicted system state.

$$\hat{Z}_k = h(\hat{X}_k^-) \tag{5}$$

where h represents the observation model used in this research. The observation model performs the transform from the world frame to the robot frame with respect to the feature information to compare the actual observations. The observation models for the corners and lamps are expressed by

$$Z = [(z_1^C)^T, \cdots, (z_n^C)^T, (z_1^L)^T, \cdots, (z_m^L)^T]^T \quad (6)$$

$$z^C = \begin{bmatrix} z_r^C \\ z_\theta^C \end{bmatrix} = \begin{bmatrix} \sqrt{(x_C - x_R)^2 + (y_C - y_R)^2} \times \dfrac{f_c}{z_C} \\ \dfrac{\pi}{2} - \tan^{-1} \dfrac{y_C - y_R}{x_C - x_R} + \theta_R \end{bmatrix} \quad (7)$$

where f_c is the focal length of the camera, $[z_r^C, z_\theta^C]^T$ denotes the radius and angle of the corner in the polar coordinates, as shown in Fig. 2. In this research, the observation model of the lamps z^L is assumed to be the same as the corners.

2.2 Update

At the update stage, the state vector and its covariance matrix P at time k are updated as follows:

$$\hat{X}_k = \hat{X}_k^- + K_k(Z_k - \hat{Z}_k) \quad (8)$$

$$P_k = (I - K_k H_k) P_k^- \quad (9)$$

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + V_k R_k V_k^T)^{-1} \quad (10)$$

where K represents the Kalman gain matrix, $H = \partial h / \partial X$ and $V = \partial h / \partial v$ are the Jacobian matrices of the observation model with respect to the state vector and the sensor noise, and R is the covariance of the sensor noise. If none of features are matched, the uncertainties of features are kept unchanged. In this case, only the robot pose is calculated by the motion model and the uncertainty of the robot pose increases.

3. FEATURE EXTRACTION

The algorithm extracts the corner and lamp features from the ceiling image. The unique features are used for the reliable data association process when indistinguishable features exist. Extracting unique corners requires a special process, since the extracted corners can include many indistinguishable corners. The lamp features are basically regarded as the unique features since they are not adjacent to each other and only a few exist in the environment.

3.1 Extraction of unique corner feature

A Harris corner detector [9] is adopted to extract corner features. Newly extracted corners are compared to the ones that were already stored in the database. For the matching algorithm, the normalized cross correlation (NCC) algorithm is adopted. NCC is the most robust correlation measure used for determining similarity between corners in two images. A search region of the indistinguishable corners is defined by the Mahalanobis distance d from the stored corner to the new corner as

$$S = H_k P_k^- H_k^T + V_k R_k V_k^T \quad (11)$$

$$d = v^T S^{-1} v < \gamma \quad (12)$$

where v is the measure of the difference between the new corner and the adjacent corners and γ is the previously determined range threshold. If an adjacent corner is found inside the search region, it is compared to the new corner by using the NCC. If any similar corners exist in the search region, NCC will give a small value which means the two corners are similar, and the new corner will be regarded as an indistinguishable corner. The corners in the database which do not turn out to be similar to the nearby corners are registered as unique corners. The indistinguishable corners are ignored in the data association process until the plane which is introduced in section 4 is estimated. An example is shown in Fig. 3. The uncertainty bound is acquired from the covariance of the corner in the EKF by projecting the covariance ellipsoid onto the image plane. Corner B in Fig. 3 is found in the search region of corner A in Fig. 3, and the comparison is made by the NCC. The proposed algorithm classifies corner A as an indistinguishable corner because it is visually similar to corner B.

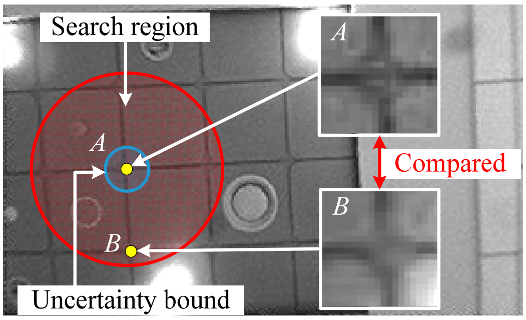


Fig. 3 Determination of indistinguishable corner.

3.2 Extraction of lamp feature

To extract a lamp as a feature, the brightest part in the image is found using the binary image. Then, each brightest part is grouped and center points are extracted by calculating the mean point of pixel points in each group.

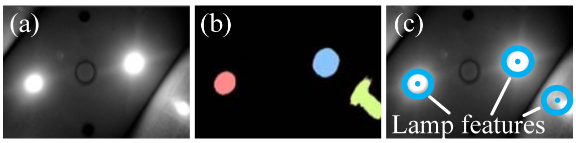


Fig. 4 Process of lamp extraction.

To reduce the computational burden, the resolution of the image is converted from 320x240 to 160x120 for the lamp extraction. Each process of the lamp extraction is presented in Fig. 4. Fig. 4(a) is the original image, the grouping result from the binary image is shown in Fig. 4(b), and the extraction result is shown in Fig. 4(c). If the lamp is reflected from the wall, it is regarded as a real lamp in the opposite side of the wall since there are no geometrical problems, as shown in Fig. 5.

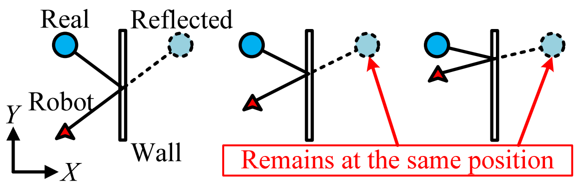
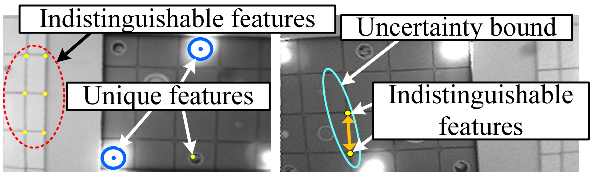


Fig. 5 Reflected lamp from wall (top view).

4. RELIABLE DATA ASSOCIATION

The false matches in the data association process occur when the uncertainty bound of a feature in the image is large. This situation can happen, for example, if indistinguishable features are close to a feature, or occlusion occurs when the robot is under the table, or the processing time is long (e.g., an embedded system for the small-size robots).

The indistinguishable features shown in Fig. 6(a) are likely to cause false matches during data association. The falsely matched features as shown in Fig. 6(b) may give incorrect coordinate information to the EKF and cause localization failure. To avoid this problem, the algorithm estimates a plane that represents the real ceiling, based on the heights of the unique corners and lamps. Then, the other features which turned out to be indistinguishable from nearby features are registered on the plane.



(a) (b)

Fig. 6 Unique and indistinguishable features.

4.1 Ceiling plane estimation

A ceiling plane is assumed to be horizontal since most of the ceilings are horizontal and only z coordinates (heights) of the unique features are needed to estimate the plane. However, the features are not always extracted from the ceiling. The features can be extracted from under the table, the wall, or the hanging decorations. Thus, the features which have small uncertainty and similar heights are grouped to estimate the ceiling plane as shown in Fig. 7, and the mean height of the features is regarded as the height of the plane.

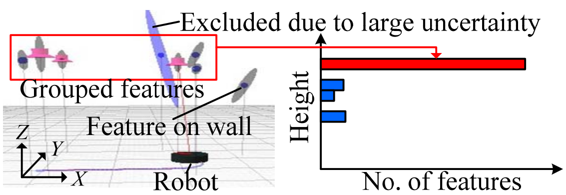
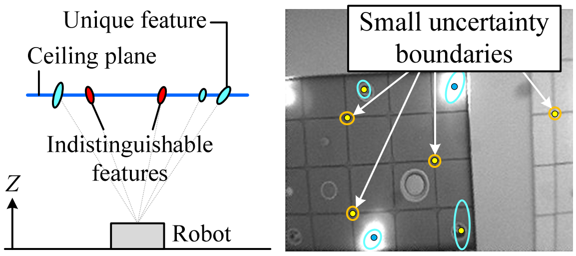


Fig. 7 Feature group selection.

4.2 Indistinguishable feature registration

The indistinguishable features are useful when the unique features are not extracted from the image due to the noise or the occlusions. Once a ceiling plane is successfully estimated, the indistinguishable features are observed for several frames. If they maintain same positions on the plane, they are registered to the database with a small uncertainty region as shown in Fig. 8. The small boundary of the uncertainty region (represented in Fig. 8(b)) reduces the probability of the false matches in the data association process significantly. If the indistinguishable features do not lie on the plane, they are ignored during the association process since they cannot be observed in the matching boundaries. For the features which are extracted from the wall, only the unique features are used for the navigation, and the indistinguishable features are ignored since they do not have any reference planes such as the ceiling plane.



(a) (b)

Fig. 8 Indistinguishable feature registration.

5. EXPERIMENTAL RESULTS

Various experiments were conducted in the real environment with the MobileRobots Pioneer 3-DX robot equipped with an upward camera having 110˚ field of view. The camera calibration was done before the experiment to define a camera matrix and distortion parameters used for the coordinate transformation and obtaining the undistorted image. The camera image was converted into the gray image with 320 x 240 pixels, and 160 x 120 pixels image was used during the lamp extraction. The size of the experimental environment was 6m x 6m which included corner and lamp features, as shown in Fig. 9.

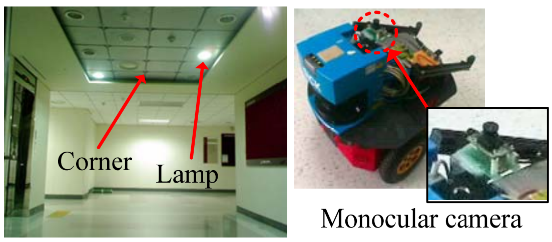


Fig. 9 Experimental environment and mobile platform.

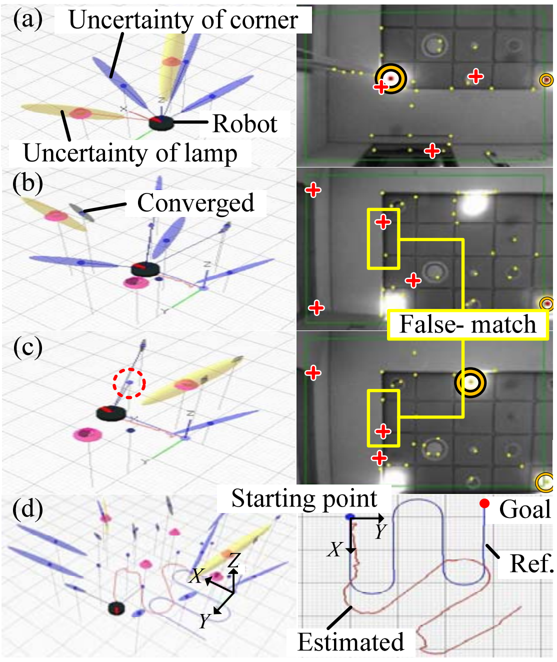


Fig. 10 Example of localization failure due to false matches.

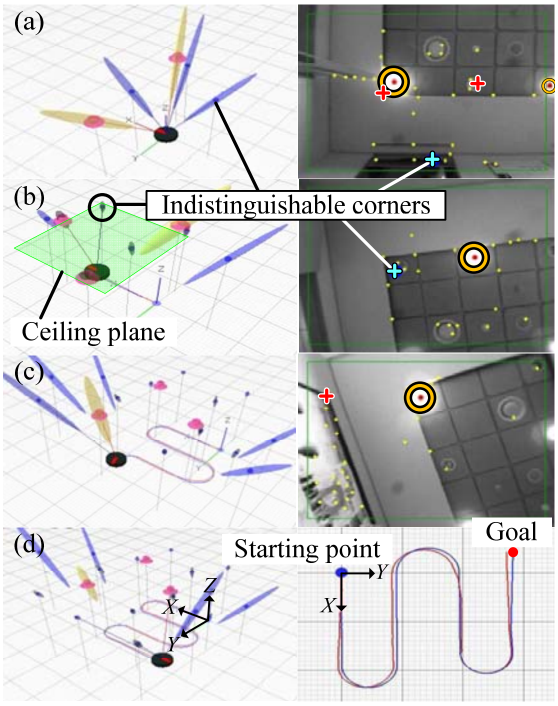


Fig. 11 Experimental results of proposed method.

An example of the localization failure due to false matches between indistinguishable features is shown in Fig. 10. The uncertainty of the corner is represented as the blue (dark) ellipsoids and uncertainty of the lamp is represented as the yellow (bright) ellipsoids. The yellow points (small circles) in the camera image are the corner points extracted from the Harris corner detector and the red points (cross-shaped) are the stored corner and lamp features in the database. The border of the lamp features is represented as the orange circles (large circles). Due to the false matches occurring between Fig. 10(b) and Fig. 10(c), the uncertainty of the false matched corner is converged to a wrong position as shown in a dotted red circle in the left side of Fig. 10(c). As a result, these false matches which occurred in the data association process caused localization failure, and this is easily confirmed by the differences between the estimated path (red) and the reference path (blue) shown in Fig. 10(d).

Fig. 11 shows the experimental result of the proposed method. At the initial state of navigation, the features which turned out to be indistinguishable are ignored before the ceiling plane is estimated as in Fig. 11(a). Once the plane is estimated, the indistinguishable features are enabled with small uncertainties as shown in Fig. 11(b). The indistinguishable corners are represented as the blue points (cross-shaped) in the image. The proposed monocular SLAM method successfully worked in the environment despite the existence of indistinguishable features within the position error of 10 cm. The unstable features with large uncertainties are removed automatically during navigation.

6. CONCLUSIONS

In this paper, a reliable data association method is proposed to prevent the localization failure by registering the features with small uncertainties on the estimated ceiling plane. The corners which have no similar corners nearby and lamps are regarded as the unique features, and the plane is estimated by the height of unique features. The EKF is used to combine the odometric and visual information to estimate the robot pose and feature positions. The validity of the proposed SLAM method is investigated by various experiments, and the following conclusions were drawn.

1. The proposed scheme enables stable navigation although many indistinguishable features exist in the environment.
2. Applying the indistinguishable features to the estimation of a robot pose makes it possible to maintain stable navigation when other features are not well extracted.
3. Since the corner and lamp features are successfully combined to estimate the robot pose, the proposed SLAM method can provide reliable performance in the various environments.

ACKNOWLEDGEMENT

This paper was performed for the Intelligent Robotics Development Program, one of the 21st Century Frontier R&D Programs funded by the Ministry of Knowledge Economy of Korea.

REFERENCES

[1] A. Davison, “Real-time simultaneous localisation

and mapping with a single camera," *Proc. of the 9th Int. Conf. on Computer Vision*, pp. 1403-1410, 2003.

[2] P. Smith, I. Reid, and A. Davison, "Real-time monocular SLAM with straight lines," *Proc. of BMVC06*, 2006.

[3] W. Y. Jeong and K. M. Lee, "CV-SLAM: a new ceiling vision-based SLAM technique," *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 3195-3200, 2005.

[4] W. Y. Jeong and K. M. Lee, "Visual SLAM with line and corner features," *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 2570-2575, 2006.

[5] M.A. Fischler and R.C. Bolles, "Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography," *Communications of the ACM*, Vol. 24, No. 6, pp. 381-395, 1981.

[6] S. Y. Hwang and J. B. Song, "Upward monocular camera based SLAM using corner and door features," *Proc. of the 17th World Congress of International Federation of Automatic Control*, pp. 1663-1668, 2008.

[7] T. Lemaire, S. Lacroix, and J. Sola, "A practical 3D bearing-only SLAM algorithm," *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 2449-2454, 2005.

[8] S. Thrun, W. Burgard and D. Fox, *Probabilistic Robotics*, MIT Press, Massachusetts, 2005.

[9] C. Harris and M. Stephens, "A combined corner and edge detector," *Proc. of the 4th Alvey Vision Conference*, pp 147-151, 1988.